CLAIMS

What is claimed is:

1.   A persistent archive of a collection of data objects tangibly embodied on a processor readable medium, the archive comprising:

5   a self-describing, infrastructure-independent representation of a logical structure for the collection; and

a self-describing, infrastructure-independent representation of the data objects.

2.   The persistent archive of claim 1 further comprising a self-describing, infrastructure-independent representation of a presentation mechanism for the data

10   objects.

3.   A method of ingesting one or more data objects into a persistent archive as claimed in claim 1, comprising:

transforming a representation of the one or more data objects into a self-describing, infrastructure-independent representation of the one or more data objects;

15   and

archiving the self-describing, infrastructure-independent representation of the one or more data objects with a self-describing, infrastructure-independent representation of the logical structure of the collection.

4.   The method of claim 3 further comprising performing the following

20   steps prior to the transforming step:

forming a self-describing, infrastructure-independent representation of a logical structure of the collection; and

forming a self-describing, infrastructure-independent representation of the data objects.

5.   A method of instantiating a persistent archive as claimed in claim 1 comprising:

retrieving from the persistent archive a self-defining representation of a logical structure for the collection;

5   creating on a medium a query-able mechanism in accordance with the logical structure;

retrieving from the persistent archive a self-describing, infrastructure-independent representation of one or more data objects; and

loading the data objects into the query-able mechanism.

10   6.   The method of claim 5 further comprising:

retrieving from the persistent archive a self-describing, infrastructure-independent representation of a presentation mechanism for the one or more data objects;

querying the query-able mechanism for one or more data objects; and

15   presenting the one or more data objects using the presentation mechanism.

7.   A method of presenting one or more data objects from a persistent archive as claimed in claim 1 comprising:

retrieving from the persistent archive a self-describing, infrastructure-independent representation of a presentation mechanism for the one or more data

20   objects;

retrieving from the persistent archive a self-describing, infrastructure-independent representation of one or more data objects; and

presenting the one or more data objects using the presentation mechanism.

8.   A method of migrating a persistent archive as claimed in claim 1, the

25   archive being maintained on a first medium, the method comprising:

retrieving the persistent archive maintained on the first medium;

65

optionally redefining the logical structure of the collection or the self-describing, infrastructure-independent representation of the one or more data objects; and

storing the persistent archive as optionally redefined in the previous step onto

5   a second medium.

9.   A processor readable medium tangibly embodying the method steps of any of claims 3-8.

10.   A system for maintaining a persistent archive as claimed in claim 1 comprising:

10   an ingestion subsystem for ingesting one or more data objects into the archive by transforming a representation of the one or more data objects into the self-defining representation of the one or more data objects, and adding the one or more transformed data objects to the archive; and

an instantiation subsystem for retrieving from the archive the self-describing,

15   infrastructure-independent representation of a logical structure for the collection, creating a query-able mechanism on a processor readable medium in accordance with the logical structure, and loading the data objects into the query-able mechanism.

11.   The system of claim 10 further comprising a migration subsystem for retrieving the persistent archive from a first medium, optionally redefining the logical

20   structure of the collection or the self-describing, infrastructure-independent representation of the one or more data objects in the collection, and storing the persistent archive as optionally redefined onto a second medium.

12.   The system of claim 10 further comprising a presentation subsystem for retrieving from the archive a self-describing, infrastructure-independent

25   presentation mechanism, retrieving from the archive one or more data objects, and

presenting the one or more data objects using the self-describing, infrastructure-independent presentation mechanism.

13.   The system of claim 10 further comprising a presentation subsystem for retrieving from the archive a self-describing, infrastructure-independent
5   presentation mechanism, querying the query-able mechanism for one or more data objects, and presenting the one or more data objects using the self-describing, infrastructure-independent presentation mechanism.

14.   The system of claim 10 wherein the instantiation system includes a plurality of drivers each configured for retrieving data from or storing data to a
10   processor readable medium.

15.   The system of claim 11 wherein the migration system includes a plurality of drivers each configured for retrieving data from or storing data to a processor readable medium.

16.   A knowledge-based persistent archive of a collection of data objects
15   tangibly embodied on a processor readable medium, the archive comprising:
        a self-describing, infrastructure-independent representation of a logical structure for the collection;
        a self-describing, infrastructure-independent representation of the data objects; and
20   a self-describing, infrastructure-independent representation of knowledge relevant to the collection.

17.   The persistent archive of claim 16 wherein the knowledge comprises relationships between concepts relevant to the collection.

18.   The persistent archive of claim 17 wherein the relationships are logical relationships.

19.   The persistent archive of claim 17 wherein the relationships are semantic relationships.

5   20.   The persistent archive of claim 17 wherein the relationships are mappings between concepts relevant to the collection and attributes of data objects.

21.   The persistent archive of claim 17 wherein the relationships are temporal relationships.

22.   The persistent archive of claim 17 wherein the relationships are
10   procedural relationships.

23.   The persistent archive of claim 22 wherein the relationships embody one or more procedures for transforming one or more data objects in the collection.

24.   The persistent archive of claim 23 wherein the relationships embody one or more procedures for transforming a representation of the one or more data
15   objects into a form ready for ingestion into the archive.

25.   The persistent archive of claim 23 wherein the relationships embody one or more procedures for transforming a representation of the one or more data objects into a form ready for instantiation onto a query-able mechanism.

26.   The persistent archive of claim 23 wherein the relationships embody
20   one or more procedures for transforming a representation of the one or more data objects into a form ready for presentation.

27.   The persistent archive of claim 17 wherein the relationships are spatial relationships.

28.   The persistent archive of claim 17 wherein the relationships are structural relationships.

5   29.   The persistent archive of claim 17 wherein the relationships embody one or more rules applicable to attributes of the data objects.

30.   The persistent archive of claim 17 wherein the relationships are algorithmic relationships between data objects and features of the data objects.

31.   The persistent archive of claim 17 wherein the relationships are
10   functional relationships between data objects and features of the data objects.

32.   A method of ingesting one or more data objects into a knowledge-based persistent archive as claimed in claim 16, comprising:

        transforming a representation of the one or more data objects into a self-describing, infrastructure-independent representation of the one or more data objects;
15   verifying the transformation of the data objects using knowledge relevant to the collection; and

        archiving the verified self-describing, infrastructure-independent representation of the one or more data objects with a self-describing, infrastructure-independent representation of a logical structure of the collection and a self-
20   describing, infrastructure-independent representation of the knowledge relevant to the collection.

33.   The method of claim 32 wherein the transforming step comprises tagging attributes of the data objects, and the verifying step comprises tagging

69

occurrences of data object attributes and their corresponding values and verifying that these occurrences are consistent with the knowledge relevant to the collection.

34.   A method of instantiating a knowledge-based persistent archive as claimed in claim 16 comprising:

5   retrieving from the persistent archive a self-defining, infrastructure-independent representation of a logical structure for the collection;

retrieving from the persistent archive a self-describing, infrastructure-independent representation of knowledge relevant to the collection;

creating on a medium a query-able mechanism in accordance with the logical

10   structure;

retrieving from the persistent archive a self-describing, infrastructure-independent representation of one or more data objects;

verifying that the one or more data objects are consistent with the knowledge relevant to the collection; and

15   loading the data objects into the query-able mechanism.

35.   The method of claim 34 further comprising:

retrieving from the persistent archive a self-describing, infrastructure-independent representation of a presentation mechanism for the one or more data objects;

20   querying the query-able mechanism for one or more data objects using the relationships between concepts relevant to the collection;

verifying that the one or more data objects are consistent with the knowledge relevant to the collection; and

presenting the one or more data objects using the presentation mechanism.

25   36.   A method of validating a knowledge-based persistent archive as claimed in claim 16, comprising:

70

retrieving from the archive a self-describing, infrastructure-independent representation of knowledge relevant to the collection; and

using the knowledge to validate the collection.

37. A method of transforming raw data records into a form capable of
5   ingestion into a knowledge-based persistent archiveas claimed in claim 16, which includes as the knowledge base a self-describing, infrastructure independent, or executable representation of a transformation procedure, comprising:

retrieving from the archive the self-describing, infrastructure independent, or executable representation of the transformation procedure;

10   executing the procedure to transform the raw records into a self-describing, infrastructure independent representation of data objects; and

adding the self-describing, infrastructure independent representation of the data objects to the archive.

38. A method of transforming a self-describing, infrastructure independent
15   representation of data objects into a form capable of instantiation onto a query-able mechanism, the data objects being from a knowledge-based persistent archive as claimed in claim 16 which includes as the knowledge base a self-describing, infrastructure independent, or executable representation of a transformation procedure, the method comprising:

20   retrieving from the archive the self-describing, infrastructure independent, or executable representation of the transformation procedure;

retrieving from the archive the self-describing, infrastructure independent representation of the data objects; and

executing the procedure to transform the self-describing, infrastructure
25   independent representation of the data objects into a form capable of instantiating onto a query-able mechanism.

71

39.   A method of transforming a self-describing, infrastructure independent representation of data objects into occurrences of attribute or element values, the data objects being from a knowledge-based persistent archive as claimed in claim 16 which includes as the knowledge base a self-describing, infrastructure independent, or

5   executable representation of a transformation procedure, the method comprising:

retrieving from the archive the self-describing, infrastructure independent, or executable representation of the transformation procedure;

retrieving from the archive the self-describing, infrastructure independent representation of the data objects; and

10   executing the procedure to transform the self-describing, infrastructure independent representation of the data objects into the occurrences of attribute or element values.

40.   The method of claim 39 further comprising using the occurrences to validate the collection.

15   41.   The method of claim 39 further comprising using the occurrences to identify exceptional conditions which are added to the knowledge base of the archive.

42.   A method of forming occurrences of attribute or element values comprising:

20   receiving data records tagged with attribute or element names; and

forming from the tagged data records occurrences of attribute or element values.

43.   The method of claim 42 further comprising using the occurrences to confirm closure of attribute or element selection for a collection formed from the

25   tagged data records.

44. The method of claim 42 further comprising using the occurrences to obtain useful information about a collection formed from the tagged data records.

45. The method of claim 44 further comprising using the occurrences to determine redundancy in a collection formed from the tagged data records.

5   46. The method of claim 42 further comprising using the occurrences to determine transformation procedures for a collection formed from the tagged data records.

47. The method of claim 42 further comprising using the occurrences to identify knowledge to be added to a knowledge base of a knowledge based persistent

10   archive formed or to be formed from the tagged data records.

48. The method of claim 47 further comprising using the occurrences to identify exceptional conditions to be added to a knowledge base of a knowledge based persistent archive formed or to be formed from the tagged data records.

49. The method of claim 42 further comprising using the occurrences to

15   check the internal consistency of a collection formed or to be formed from the tagged data records.

50. The method of claim 42 further comprising transforming the occurrences into an inverted attribute index.

51. The method of claim 42 further comprising transforming the

20   occurrences into tagged data records.

52. The method of claim 42 further comprising transforming the occurrences into a form capable of being ingested into a persistent archive.

53.   The method of claim 42 further comprising transforming the occurrences into a form capable of being instantiated onto a query-able mechanism.

54.   The methods of any of claims 32-53 tangibly embodied on a processor readable medium.

5   55.   A knowledge-based persistent archive of a collection of data objects tangibly embodied on a processor readable medium comprising:

at least one representation of the collection of data objects;

at least one self-describing, infrastructure-independent or executable specification of one or more transformations relating to the collection; and

10   at least one self-describing, infrastructure-independent or executable specification of one or more rules encoding knowledge relevant to the collection.

56.   The archive of claim 55 wherein one of the representations of the collection is a self-describing, infrastructure-independent representation.

57.   The archive of claim 55 wherein one of the representations of the collection is raw data.

15

58.   The archive of claim 55 wherein one of the representations of the collection is capable of presentation.

59.   The archive of claim 55 wherein one of the representations of the collection is capable of instantiation onto a query-able mechanism.

20   60.   The archive of claim 55 wherein one of the representations comprises occurrences of attribute or element values.

61.   The archive of claim 55 wherein one of the representations comprises one or more inverted attribute indices.

62.   The archive of claim 55 wherein one of the representations comprises a topic map.

63.   The archive of claim 55 wherein one of the representations is capable of migration onto another medium.

5   64.   The archive of claim 55 wherein one of the transformations is content-preserving.

65.   The archive of claim 64 wherein one of the transformations is invertible.

66.   The archive of claim 55 wherein one of the transformations is

10   configured to produce data objects in a form suitable for ingestion into the archive.

67.   The archive of claim 55 wherein one of the transformations is configured to produce data objects in a form suitable for instantiation onto a query-able mechanism.

68.   The archive of claim 55 wherein one of the transformations is

15   configured to produce data objects in a form suitable for presentation.

69.   The archive of claim 55 wherein one of the transformations is configured to produce data objects in a form suitable for migration.

70.   The archive of claim 55 wherein one of the transformations is configured to produce occurrences of attribute or element values.

20   71.   The archive of claim 55 wherein one of the transformations is configured to produce one or more inverted attribute indices.

72.   The archive of claim 55 wherein one of the representations of the collection is a product of one of the transformations.

73.   The archive of claim 55 wherein one of the representations of the collection is an input to one of the transformations.

5   74.   A method of automatically placing one or more data objects from a persistent archive as claimed in claim 55 into a form suitable for instantiation onto a query-able mechanism comprising:

retrieving from the archive a self-describing, infrastructure-independent or executable specification of one or more transformations relevant to the collection;

10   retrieving from the archive a representation of one or more data objects in the collection; and

executing the specification to automatically place the one or more data objects into a form suitable for instantiation onto the query-able mechanism.

75.   A method of automatically validating a collection of data objects

15   within a persistent archive as claimed in claim 55 comprising:

retrieving from the archive a self-describing, infrastructure-independent or executable specification of one or more rules relevant to the collection; and

executing the specification to automatically validate the collection.

76.   The method of claim 75 further comprising validating the collection by

20   performing the following substeps:

producing occurrences of attribute or element values; and

determining that the occurrences are consistent with the rules encoded by the specification and any valid exceptions.

77.   A method of automatically presenting one or more data objects from a

25   persistent archive as claimed in claim 55 comprising:

76

retrieving from the archive a self-describing, infrastructure-independent or executable specification of one or more transformations relevant to the collection;

retrieving from the archive a representation of one or more data objects in the collection; and

5   executing the specification to automatically place the one or more data objects from the collection in a form suitable for presentation.

78.   A method of automatically placing a persistent archive as claimed in claim 55 into a form suitable for migration to a new medium comprising:

retrieving from the archive a self-describing, infrastructure-independent or

10 executable specification of one or more transformations relevant to the collection; and

executing the specification to automatically place the collection into a form suitable for migration to a new medium.

79.   The system of claim 10 further comprising an engine for executing self-describing, infrastructure-independent, or executable specifications.

15   80.   The system of claim 79 further comprising a validation subsystem for validating the collection by commanding the engine to execute at least one self-describing, infrastructure-independent or executable specification encoding one or more rules relevant to the collection.

81.   The system of claim 79 further comprising a transformation subsystem

20 for transforming one or more data objects in the collection by commanding the engine to execute at least one self-describing, infrastructure-independent, or executable specification of one or more transformations relevant to the collection.

82.   The methods of any of claims 74-78 tangibly embodied on a processor-readable medium.

83.   A persistent archive of a collection of data objects tangibly embodied on a processor-readable medium, the collection having a logical structure, comprising:

first means for representing the logical structure of the collection; and

second means for representing the data objects in the collection.

5   84.   The persistent archive of claim 83 further comprising third means for representing knowledge relevant to the collection.

85.   A persistent archive of a collection of data objects tangibly embodied on a processor-readable medium comprising:

first means for representing the data objects or the collection;

10   second means for specifying one or more transformations relating to the collection; and

third means for specifying one or more rules relating to the collection.